DOCUMENT RESUME

ED 422 049 JC 980 372

AUTHOR	Petrowsky, Michael C.
TITLE	Using a Multiple Choice Exam in Economics to Satisfy Possible Institutional Effectiveness Requirements: Some Considerations.
INSTITUTION	Glendale Community Coll., AZ.
PUB DATE	1998-08-00
NOTE	24p.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Community Colleges; *Economics Education; Guessing (Tests); Institutional Evaluation; *Macroeconomics; *Multiple Choice Tests; Objective Tests; Tests; Two Year Colleges
IDENTIFIERS	Glendale Community College AZ

ABSTRACT

This essay describes the problems involved in developing a comprehensive multiple choice exam in macroeconomics. It also attempts to promote dialogue about the testing instruments used to satisfy institutional effectiveness requirements. One section briefly compares the advantages and disadvantages of using essays instead of multiple choice exams. Another section explores general rules for constructing multiple choice exams, suggesting 22 things to do and avoid in order to improve test quality. Attention also is paid to the special problems faced in forming pre- and post-test multiple choice exams, specifically for use in the macroeconomics course at Glendale Community College (AZ). Such problems involved in test construction include time constraints, selecting a testing instrument, and choosing question type by cognitive and content categories. (AS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 422 049

USING A MULTIPLE CHOICE EXAM IN ECONOMICS TO SATISFY POSSIBLE INSTITUTIONAL EFFECTIVENESS REQUIREMENTS: SOME CONSIDERATIONS

Social Science Department

Glendale Community College
August, 1998

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
M. Petrowsky
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

JC 980 372

GLENDALE
COMMUNITY
COLLEGE

6000 W. OLIVE AVE.
GLENDALE, AZ 85302
(602) 435-3000

DATE: August 15, 1998

TO: Paul DePippo

FROM: Michael C. Petrowsky

RE: Developing Multiple Choice Exams in Order to Satisfy Possible IE/North Central Requirements

Enclosed is a report that describes the numerous problems that are involved in developing a comprehensive multiple choice exam in macroeconomics. In addition to the narrative, I have also enclosed relevant literature as background material. The appendices include sample exams as well as well as a possible procedure that we could use to produce departmental finals.

This document should be thought of as a "working draft" that is designed to stimulate discussion and perhaps prepare us in the event that we have to create standardized exams in order to satisfy North Central requirements.

MARICOPA COMMUNITY COLLEGES

USING A MULTIPLE CHOICE EXAM IN ECONOMICS TO SATISFY POSSIBLE INSTITUTIONAL EFFECTIVENESS REQUIREMENTS: SOME CONSIDERATIONS

Social Science Department

Glendale Community College
August, 1998

CONTENTS

USING A MULTIPLE CHOICE EXAM IN ECONOMICS TO SATISFY POSSIBLE INSTITUTIONAL EFFECTIVENESS REQUIREMENTS: SOME CONSIDERATIONS

Transmittal 1

I. INTRODUCTION 1

II. ESSAY VERSUS MULTIPLE CHOICE 2

III. GENERAL CONSIDERATIONS IN USING MULTIPLE CHOICE TESTS 4

IV. SPECIAL CONSIDERATIONS IN DESIGNING AND USING MULTIPLE CHOICE TESTS IN ECONOMICS AT GCC 7

V. AN ARRAY OF TEST OPTIONS 17

APPENDICES:

A. Multiple Choice Versus Essays in Economics
B. Journal of Educational Measurement Article
C. Literature Search on Multiple Choice
D. Excerpts from two books
E. ECN 11 Course Competencies
F. Group Procedure for Selecting Multiple Choice Questions
G. TUCE
H. Randomly selected exam of 66 questions
I. Randomly selected exam of 33 questions
J. Randomly selected exam of 60 questions
K. Randomly selected exam of 30 questions
L. Competency based exam of 33 questions
M. Competency based exam of 60 questions

I. INTRODUCTION

The purpose of this essay is to promote dialogue about the testing instruments that we may need to use in order to satisfy institutional effectiveness (IE) requirements. While we have received little guidance as to what is needed or expected from North Central, it does seem prudent to begin at least thinking about what we would do if discipline based testing instruments become the required standard practice.

In the first section, I briefly compare and contrast the advantages and disadvantages of using essays versus multiple choice exams for the purpose of complying with institutional effectiveness mandates. For a number of very good reasons (time, reliability, etc.,) it seems beneficial to use the multiple choice format, although I have provided some information which could generate essay questions as well.

The next section discusses, in a general way, the "do's" and "don'ts" involved in constructing multiple choice exams. While it is somewhat improbable that we would be constructing our own exams from scratch, some of this material may be helpful in spotting "bad" questions that come from the numerous published test banks. (Or at least that is my hope.)

After these general considerations have been discussed, attention is then paid to the special problems that are involved in constructing pre and post test multiple choice exams for use in the macroeconomics course at GCC. Although no solutions are offered, it is hoped that a discussion of these problems will (heroically) lead to more effective decision making in terms of test construction and selection.

Finally, the last section outlines a possible procedure that we could use to produce departmental exams. This is supplemented by the development of "sample" exams that I generated using the test bank software that accompanied the Gregory and Ruffin text. Again, these exams are for illustrative purposes only, and are meant to stimulate discussion.

II. ESSAY VERSUS MULTIPLE CHOICE

From the literature I have scanned, there seems to be numerous advantages to using a multiple choice test. But the biggest advantage, perhaps, is that the multiple choice format allows for more objective assessment because it eliminates bias in scoring. From an IE perspective, this is a definite plus, and especially so when contrasted with the serious reliability problems that are part and parcel of the essay test instrument.

Second, a multiple choice exam enables an instructor to test for a larger body of knowledge. This attribute becomes especially important if our department moves to a comprehensive final exam. Essay tests, by contrast, are limited , by definition, to only that material which the essay(s) cover.

For multiple choice tests, then, the twin advantages of higher reliability and greater content coverage provide a net plus over the essay format. These advantages are further magnified when one adds in the benefits of statistical analysis and ease of

administration. These advantages, moreover, also apply to other selected response questions such as those found in matching type exercises.

In Appendix A I have enclosed two chapters from the book, *Teaching Undergraduate Economics: A Handbook for Instructors*, by William Walstad and Phillip Saunders. The first chapter (by Walstad) sings the praises of multiple choice tests, while the other chapter (by Saunders and Arthur Welsh) discusses the merits of the essay format. Both chapters are highly readable and quite informative.

The evidence as to which test type is best in terms of measuring student achievement is unclear and sketchy. In Appendix B I have enclosed a fairly recent article from an ED Journal that gives a slight edge to the multiple choice supporters. I have also enclosed (Appendix C) a literature search I performed on the use of multiple choice tests in economics and other subjects. Again, the results appear to be rather inconclusive.

In developing a comprehensive final exam, my preference is to use the multiple choice format. But I also think that it is possible to generate essay questions that minimize some of the reliability problems previously noted. For example, the 13th edition of the McConnell and Brue textbook comes with a third test bank (the first two contain multiple choice questions) that contains 946 constructed response questions. In this total, there are 690 essay questions with sample answers. Using something like this could mitigate some of the scoring problems with essay questions that were previously discussed. In addition to the essay questions, this test bank also contains 256 problems, which could generate any number of problem based exams which have the virtue of being

objective in nature. Given all this, it may be possible to develop a final exam which has multiple choice, essays, and problems. While this hybrid is not discussed here, it should not be precluded as an option.

II. GENERAL CONSIDERATIONS IN USING MULTIPLE CHOICE TESTS

In the educational literature, a distinction is made between selected response questions and constructed response questions. The selected response format includes multiple choice, matching, and true-false questions. The name "selected response" derives from the fact that the student has to select the answer. In the constructed response format, the student has to construct or develop an answer. Examples of this format include essay questions, performance based testing (playing a musical instrument, typing, etc.) and product based testing such as the development of an art portfolio.

From the literature, constructed response formats are generally used for measuring skill, while both the selected response and constructed response formats are used when measuring knowledge. Because the study of economics involves knowledge that utilizes critical thinking and problem solving, multiple choice questions can be used, although there are numerous obstacles that need to be avoided if the virtues of multiple choice test usage are to be maximized.

Following this, I found two books that were quite helpful in assessing the pitfalls that are involved in constructing multiple choice tests. Relevant excerpts from these books are enclosed as Appendix D. They are: *Educational Measurement and Testing*, by William Wiersma and Stephen Jurs; and, *Developing and Validating Multiple Choice Test Items*, by Thomas M. Haladyna. In what follows, I have summarized what I have gleaned from these tomes.

In the jargon of education, a multiple choice question is broken up into two parts. The first part is the **stem**, which is the lead statement that provides the necessary information and introduces the question. The second part is the **options** section, which consists of the correct answer and **distractors** (questionably wrong answers). In designing a multiple choice question, the two books emphasized a number of things to do and/or avoid in order to improve the quality of multiple choice questions. These included:

1. Do not indicate correct answers by the grammar.
2. Make all distractors the same length.
3. Make all distractors plausible.
4. Avoid repeating key words in the stem and the options.
5. Avoid trick items.
6. Avoid over specific knowledge.
7. Keep vocabulary consistent.
8. Minimize examinee reading time of questions.
9. Keep test questions independent of each other.

10. Base each item on an important learner outcome.
11. State the item in a question format instead of a completion format.
12. Avoid excess verbiage in the stem.
13. Word the stem positively; avoid negative phrasing.
14. Avoid items based on opinion.
15. Include the central idea and most of the phrasing in the stem.
16. Ensure that directions in the stem are clear and that the wording lets the examinee know exactly what is being asked.
17. Place options in logical or numerical order.
18. Options should not be overlapping.
19. Avoid "all of the above."
20. Avoid "none of the above."
21. Phrase options positively.
22. Avoid the use of words ("specific determiners") such as "always," "never," "totally," "absolutely," and "completely."

Needless to say, a great deal of this is obvious, but guidelines like this may still be of help if we decide to construct our own questions or if we decide to select multiple choice questions from a published test bank. As regards the latter, the use of guidelines may be especially helpful given the uneven quality of questions that are found in much of the published material.

IV. SPECIAL CONSIDERATIONS IN DESIGNING AND USING MULTIPLE CHOICE TESTS IN ECONOMICS AT GCC

Aside from the general problems associated with the construction of multiple choice tests, there are a number of other factors which need to be considered which are specific to GCC and to the field of economics. These difficulties are discussed in the following section.

1. Time Constraints: Pre and Post Test

Time is a physical constraint that affects the breadth of the testing instrument. Our classes run in 50 minute, 75 minute, and 150 minute intervals. We also have two hour final exams. If we elect to adopt a standard final exam, we could utilize a fairly comprehensive testing instrument that has anywhere from 60 to 75 multiple choice questions. (Assuming that a reasonable multiple choice question takes slightly less than two minutes to answer.) On the other hand, if we move in the direction of a pretest and posttest format, the test instrument will have to be able to accommodate a 50 minute class period. This means that the number of multiple choice questions on the exam will be greatly diminished, with an upper limit of 30 to 35 questions. The exam's coverage of content will be accordingly less.

The decision, then, to either have a final exam, or to adopt a pre and posttest format, must take into account this time factor. The use of a pretest and posttest does

have statistical advantages in terms of measuring outcomes. But there are disadvantages which need to be considered as well. The pre and posttest format will be less comprehensive than its two hour counterpart. This may make Department consensus on the types of questions which are to be included on the exam harder to obtain than for a two hour exam which has a greater number of questions on which to accommodate individual faculty preferences.

Finally, two other considerations come to mind. First, the pre and posttest format will have to use up a classroom hour of instruction at the beginning of the semester. Aside from this opportunity cost, there is also a potential (scare) effect on retention which is hard to measure. Second, the pre and posttest format would presumably dictate a 50 minute exam which would be used in our two hour final exam slot. Whether this is procedurally legal is uncertain at best.

2. Selection of a Testing Instrument

A variety of testing instruments can be used, and these all carry with them advantages and disadvantages. For a multiple choice format, the test instrument options consist of (a) using the standardized Test of Understanding in College Economics (TUCE), designing a test from a published test bank that is based on the text we use, and (3) designing a customized test that might include test questions from a number of published test banks as well from sources (such as us) that are not published. Each of these options is discussed in the section that follows.

A.Using the Test of Understanding in College Economics (TUCE). There are numerous advantages to using the TUCE as a testing instrument. First, it is norm referenced. Second, it is accompanied by an instructor's manual that contains a wealth of past information on how the test was constructed and used. And, finally, it is flexible enough to be used as a pre and post test, for it contains between 30 to 33 questions. (The last three questions are international trade and finance questions that can be used in either the macro or micro versions of the exam.)

But there are disadvantages to using the TUCE. A great many of the questions are policy oriented. This makes its use difficult for those instructors that favor mechanics or positive economic concerns over normative elements. Because the TUCE is textbook neutral (which from one perspective is an advantage), it has a generic quality which may prove troublesome for many of our students who are textbook dependent. (And, it should be pointed out, the TUCE's norms have been largely developed by using four year college and university students; very few community colleges were included in the data base that generated subsequent norms.)

While the TUCE can be easily used as a pre and posttest, its use as a comprehensive final exam remains problematic at best, for it is limited to its 33 question, 50 minute format. Additional questions, of course, could be added, but they would have to be of non TUCE origin. What this would do to the norm referenced qualities (and advantages) of the TUCE is not known.

B. From a published test bank that is based on an assigned text. Just about every textbook carries with it ancillary materials that include a test bank of multiple choice and true – false questions. These test banks, however, vary in terms of quality (good, clear questions versus unclear, ill framed, etc.) as well as in accessibility or ease of use. From my experience, the best test banks, in terms of clear, concise questions, were those that accompanied the Spencer and McConnell texts, while the worst test banks were associated with some texts that we have used in recent years.

In addition to question quality, ease of use is also another consideration. Most test banks come with computer software, but this software also varies in terms of options, simplicity of usage, and accuracy. In addition, some publishers have 800 numbers which enable a user to "call in" questions which are then printed out and mailed back to the faculty member. This latter option has the potential for generating customized exams (that are based on the assigned text) even if the computer test software is either not working (which sometimes happens) or is simply not available.

Aside from these overall concerns, a major decision will arise over whether to base our exams on either (1) a test bank that is tied to the textbook that we are using, or (2) to make the exams up from a variety of test banks and other sources. If this latter option is followed, then any subsequently generated exams would be text book neutral.

There are numerous advantages to generating exams using a test bank that is based on the text book that we have assigned to our students. First, there is the question of terminology, emphasis, and even syntax. Different textbooks define terms in slightly different ways and even use slightly different symbolic notation. (As an example, I have seen the short run aggregate supply curve denoted by SRAS, AS, and ASSR.) This can create problems for our students if we use a test bank from a text book that we are not using and that test bank uses slightly different notations and definitions in its test question structure.

Second, there is also the comfort level associated with using a test bank that is keyed with an assigned text. Students (presumably) will have read the text book and will thus be better prepared for language nuances that will be on the exam. This may also have positive effects on student retention by reducing student fear and uncertainty. Faculty, too, will be more sensitive to questions that (inevitably) need to be replaced on the exam because of wordiness, ambiguity, item analysis, and the like. In sum, there is a comfort level generated here that should not be discounted.

Still, a text book dependent exam does have its downsides. For one, we would be limited to only those questions that are found in the test bank. Given the quality and range of some test banks, this could prove troublesome. Second, frequent textbook changes would also require that we change test banks as well. Thus, adding to the onerous (and some would say unpleasant) task of textbook selection would be the equally unpleasant job of generating new exams. Finally, frequent text book changes, and the

subsequent need to generate new exams, might preclude the historical longitudinal analysis of data over the years because of test non comparability. As an example, we would lose the ability to compare the effectiveness of different textbooks if we used different exams.

C. Using a customized test. This last option involves the formulation and creation of a multiple choice test that uses new (original) questions as well as test questions from a variety of test banks. The use of such a customized exam would eliminate many of the disadvantages of the text book dependent exam that were discussed in the previous section. Yet there are many disadvantages to this approach as well. Test construction would be much more difficult and would have to take into account the differences in terminology and notation that are found in many textbooks. Then, too, faculty consensus might become more complex and uncertain. Overall, this approach would be more time consuming in the short run, but more time efficient in the long run if faculty consensus could be easily reached and if a textbook neutral exam could be constructed that took into account the terminological and notational issues previously described.

3. Choosing Question Type By Cognitive Category

The previous section discussed the decisions that will have to be made concerning the selection of a testing instrument. But decisions will also have to be made concerning

the *types* of questions that will be used. For purposes of exposition, this is broken down by cognitive and content categories. In what follows, an attempt is made to lay down a bare outline of these considerations.

Many test banks have developed a matrix that places each multiple choice question into one or more cognitive categories. As an illustration, these categories are sometimes labeled "definitional," "conceptual," and "problem/application." Definitional questions usually involve the recognition of a term or concept, while conceptual questions typically have an understanding component that goes beyond simple recognition. Problem/application questions frequently involve the manipulation of a graph or a set of numbers in order to arrive at the correct solution. In addition to this, there is sometimes superimposed on this cognitive classification a judgment as to whether the question is "easy," "moderately difficult," or "hard." As a result, the test bank might have a matrix of nine possibilities that range from "easy/definitional to "hard/problem-application."

For our purposes, a decision may have to be made as to what type of "cognitive mix" we desire. Many computerized test banks give you this option, so it becomes relatively easy to reach the mix that you want. Some even include a random setting that will automatically do it for you. Whatever the outcome, though, it is clear that we should at least think about this in a decisive way before generating exams.

As with any decision, there will probably have to be some clear criteria that drive the issue of question selection. One could argue, for example, that the exam should be as

easy as possible so as to minimize student frustration. Or, alternatively, that the exam should be as hard as possible and also emphasize problems/applications over memorization so as to promote long term learning. The possibilities here, while endless, only illustrate the idea that some criteria should be expressed. This becomes especially important if we use the Gregory/Ruffin text, for the test bank associated with this book does not specify question type or difficulty.

4. Choosing Question Type By Content Category

Another decision that will have to be made concerns the selection of questions on the basis of content. That is, what content/material do we want emphasized in these exams? The three options here involve selecting questions on the basis of either (1) course competencies, (2) making a selection that is random, or (3) making a nonrandom selection that is (presumably) based on specific chapters from one or more texts. Each option is discussed in the next section.

A. Choosing questions by course competency. Our macroeconomic principles course has an official course description as well as 13 official course competencies. (See Appendix D.) To a certain extent, these competencies can guide the selection of test questions, although a number of uncertainties are generated in the process. For example, should a test cover all 13 competencies or only a portion (say 80%) of them? Related to

this is also the question of weighting. Does each competency receive an equal number of questions? Or is the material in the latter half of the course given more weight?

These mechanical considerations aside, other difficulties remain. As an example, while a given competency can guide the selection of questions, the correspondence is by no means clear cut, for there may be many, many test bank questions for a given competency. Some of these questions may be clearly related to the competency, while others are less so. The competencies, then, are not a cookbook but rather a set of general guidelines.

Other problems are perhaps more substantive. The question of where to place international issues is still largely unresolved, with the result that many colleges and universities would be just as likely to place some of these international issues (as exemplified by competencies 12 and 13) in their microeconomics course.

Finally, the dated nature of the course competencies may also pose problems vis-à-vis text book selection. Current textbook trends appear to be de emphasizing Keynesian economics and re emphasizing problems of economic growth. Some texts are even downplaying short run aggregate supply as a topic. Because of this, future text books may be increasingly out of joint with current course competencies. This would ultimately involve an overhaul of both the course competencies and any subsequent tests based on these competencies.

B. Choosing questions randomly. Most test bank software will enable the user to randomly pull questions from a given chapter. If this approach were followed, it would require that we first agree on what chapters from the assigned text were to be covered. As an example, the Ruffin and Gregory text (hardbound) has approximately 16 chapters in its macro portion. These include the introductory chapters (1-4) as well as the specific macro components that are in chapters 23-34. (International issues are treated in chapters 35-37.) The variations here, in terms of chapter selection, could be the following:

- *Comprehensive*. Include all chapters, that is chapters 1-4, 23-37, for a total of 19 chapters.
- *Delete international*. Chapters covered would be 1-4, 23-34, for a total of 16 chapters.
- *Emphasize macro*. Covered chapters would include 23-34, for a total of 12 chapters.
- Other combinations.

Once chapters were selected, the computer could then randomly select the agreed upon number of questions. The advantages to this approach are that (1) it is easy and that (2) it removes the thorny problem of personal selection. The disadvantage, of course, is that there is no longer a correspondence with course competencies. Random selection may also generate questions that have little or no appeal to faculty.

C. Choosing questions non randomly. This option is similar to the above with the

exception that questions from each chapter are selected by faculty using either some predetermined criteria or are simply based on personal judgement. An example of a predetermined selection criteria would be to avoid any content material that appears in appendices, with the result that questions from these appendices would not be used. Similar rules could be used to include/exclude particular topics or chapters. The article by Walstad (Appendix A) discusses the use of a "test log" that may be of help here.

In the end, however, this approach would still have to largely rely on personal judgment. While this appears arbitrary, it may in fact be the best approach because of its acceptance by faculty and by the fact that question selection would probably be improved by the elimination of inappropriate questions.

V. AN ARRAY OF TEST OPTIONS

The test options below are for illustrative purposes only and are designed to stimulate discussion.

- Appendix F suggests procedures that could facilitate group decision making on test question selection.
- Appendix G is a copy of the TUCE.
- Appendix H is a 66 question exam, randomly selected, from the Ruffin & Gregory text. Designed as a comprehensive final exam.

International questions (61-66) are at end of exam and cover chapters 35-37.

- Appendix I is a 33 question exam, randomly selected, from the Ruffin & Gregory text. Designed as a pre and posttest. International questions (31-33) are at the end of the exam.
- Appendix J is a 60 question exam, randomly selected, from the Ruffin & Gregory text. International material is not included; covered chapters are 1-4, 23-34.
- Appendix K is a 30 question exam, randomly selected, from the Ruffin & Gregory text. Questions are from chapters 1-4, 23-34. International issues are not included.
- Appendix L is a 33 question exam, loosely based on competencies, from the Ruffin & Gregory text. 30 questions are from chapters 1-4, 23-34; questions 31-33 are from chapters 35-36 (international).
- Appendix M is a 60 question exam, loosely based on competencies, from the Ruffin & Gregory text. 54 questions are from chapters 1-4, 23-34; questions 55-60 are from chapters 35-36 (international).

Note: Appendix contains copyrighted materials and are not included for ERIC use.